

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 04-299410

(43)Date of publication of application : 22.10.1992

(51)Int.Cl.

G06F 3/16

G06F 3/16

G10L 3/00

G10L 3/02

(21)Application number : 03-063444

(71)Applicant : NIPPON TELEGR & TELEPH CORP  
<NTT>

(22)Date of filing : 27.03.1991

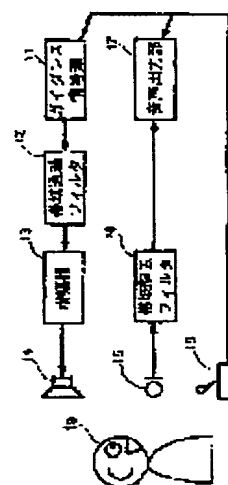
(72)Inventor : NAKADAI YOSHIO  
SUZUKI YOSHITAKE

## (54) VOICE INPUT DEVICE WITH GUIDANCE VOICE

### (57)Abstract:

PURPOSE: To extract only a voice signal that does not include the guidance voice component when the speaker's voices are superposed on the guidance voices.

CONSTITUTION: The guidance signal outputted from a guidance signal source 11 is transmitted through a BPF 12, amplified by an amplifier 13, and outputted through a speaker 14. Then the amplified guidance signal is received by a speaker 19. Meanwhile the band stop signal of the BPF 12 is extracted out of the voice signal of the speaker 19 received through a microphone 15. Thus it is possible to obtain a voice signal that does not include the guidance signal component even though the guidance voice is superposed on the voice signal of the speaker 19 in terms of time.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-299410

(43) 公開日 平成4年(1992)10月22日

(51) Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/16	3 4 0 A	8323-5B		
	3 3 0 E	8323-5B		
G 1 0 L 3/00		R 8946-5H		
3/02	3 0 1 C	8842-5H		

審査請求 未請求 請求項の数1(全 6 頁)

(21) 出願番号 特願平3-63444

(22) 出願日 平成3年(1991)3月27日

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72) 発明者 中▲台▼ 芳夫

東京都千代田区内幸町一丁目1番6号 日

本電信電話株式会社内

(72) 発明者 鈴木 義武

東京都千代田区内幸町一丁目1番6号 日

本電信電話株式会社内

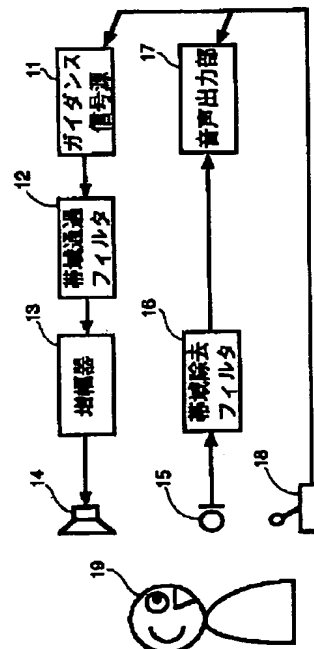
(74) 代理人 弁理士 三好 秀和 (外1名)

(54) 【発明の名称】 ガイダンス音声付き音声入力装置

(57) 【要約】

【目的】 ガイダンス音声に発声者の発声が重畳されてもガイダンス音声の成分を含まない音声信号のみを取り出す。

【構成】 ガイダンス信号源11から出力されるガイダンス信号を帯域通過フィルタ12に通した後、増幅器13によって増幅してスピーカ14から出力して発声者19に受聴させ、マイクロホン15で受信した発声者19の発声信号から前記帯域通過フィルタ12の阻止帯域の信号を取り出し、ガイダンス音声と発声者の音声信号とが時間的に重畳しても、ガイダンス信号の成分を含まない音声信号を得ることができる。



## 【特許請求の範囲】

【請求項1】 発声者に対して発声を促すガイダンス信号を生成する信号生成手段と、該信号生成手段の出力信号から特定の周波数帯域の信号を取り出す帯域通過フィルタと、該帯域通過フィルタの出力信号を発声者に受聴させるべく該出力信号を音声信号として発生する拡声器と、前記音声信号にตอบสนองして発声者から出力される音声を受信し、電気信号として出力する信号入力手段と、該信号入力手段の出力信号から前記特定の周波数帯域以外  
10 の信号を取り出す帯域除去フィルタとを有することを特徴とするガイダンス音声付き音声入力装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、ガイダンス音声を発声者に対して出力して発声者から発声を促し、この発声者からの発声を受信するガイダンス音声付き音声入力装置に関する。

## 【0002】

【従来の技術】発声者から音声入力装置を離して使用する、いわゆるハンズフリーで使用する音声認識装置などで、装置に音声を入力する発声者に対して、発声のタイミングを明らかにするためのガイダンス信号発生装置などを付加することは、認識装置の稼働時間やアルゴリズム上の負担を軽減し、かつ、発声者に対しても装置に対する親和感を与える上で重要である。このようなガイダンス信号は、一般に音声認識装置がキーボードなどの手で入力する装置の代用として使われるために、また発声者自身が発声結果をフィードバックして即時に理解することが望まれるために、視覚や触覚に訴える装置よりも、聴覚に訴える装置が望まれる。すなわち、ガイダンス信号発生装置として、一般には、音声によるガイダンス方法が望ましい形である。

【0003】ところが、音声によるガイダンス方法では、ガイダンス音声自体が発声者の出す音声と同一の音声周波数帯域の信号であるため、ガイダンス音声の生成途中で発声者が音声を装置に向かって入力するような場合を考えると、音声信号入力器では発声者の音声と一緒にガイダンス音声を受信するため、装置にはガイダンス音声の重畳した音声パターンが入力されることになり、認識装置において認識誤りを生じる原因となる。

【0004】この問題を解決する方法は4つ考えられる。第1には、ガイダンス音声の発生途中での音声入力を受け付けられない方法がある。しかし、この方法ではガイダンス音声の発生途中まで発声者に発声させないよう時間的制約を課すこととなる。また、ガイダンス音声の発生途中で発声者が発声した場合、ガイダンス発生時間中の入力音声を分析することができないため、いわゆる語頭切れという現象を生じ、認識装置での認識誤りの原因となる。第2には、ガイダンス音声の音量を低く抑える方法がある。この方法には、音声入力装置のアルゴ  
50

リズムによって発声者の音声を検知した場合にガイダンス音声を打ち切る方法も含まれる。しかし、高騒音下で使用される認識装置では、ガイダンス音声の音量を小さくすることは明瞭なガイダンス音声を与える上では実用的ではない。また、音声を検知した場合にガイダンス音声を打ち切る方法は、音声検出のアルゴリズムが完全でなければ第1の方法で示したものと同じ語頭切れを生じ、また、残響のあるような部屋で装置を使用する場合ではガイダンス音声の残響成分までを完全に打ち消すことは難しい技術である。第3には、ガイダンス音声を発声するスピーカーを音声入力装置から音響的に切り離す方法がある。この方法は、例えば、電話機の実話器のように、発声者にガイダンスを受聴させる音響系と、発声者の音声を入力させる音響系とを分離するものである。しかし、この方法では、音声入力音響系との分離を図るため発声者に密着したガイダンス発生装置が必要となり、発声者に装置を装着する煩わしさを与える。第4には、装置に入力された音声信号から重畳しているガイダンス音声を事後的に除去しようとする方法がある。この例としては、適応予測フィルタの手法により、装置内部にガイダンスの原信号を予め記憶しておき、音声入力装置で受信した音声信号からガイダンス音声信号とそれに付随して発生する音響信号を原信号に基づいて推定し除去しようとする方法がある。しかしながら、この方法では、ガイダンス信号の推定が完全でなかった場合、発声者の音声信号に対して波形歪を生じさせ、認識装置での認識率の劣化につながる。さらに、適応予測フィルタの係数の学習のために、発声者に発声を促すガイダンスとは別にフィルタの学習のためのガイダンス信号と学習時間を必要とする。従って、列記したこれら4つの方法では、いずれも認識装置に悪影響を与えず、かつ、発声者に負担とならない形でのガイダンス音声と入力音声との分離を図ることは難しい。

## 【0005】

【発明が解決しようとする課題】上述した各方法は、それぞれ欠点があり、ガイダンス音声に発声者の発声が重畳された場合には、本音声入力装置を認識装置に適用した場合に、誤認識を生じるという問題がある。

【0006】ところで、上述した各方法のうち、第4の方法は、適応予測フィルタによる事前の学習が必要であり、また、フィルタの推定が完全でない場合に発声者の音声自体に歪を生じるという点で問題があることを指摘した。しかし、この問題は、除去しようとするガイダンス音声の周波数帯域を限定し、その帯域内で除去を行うことにより、その他の周波数帯域における音声波形に歪を与えずに解決できる。また、この除去方法においては、適応予測フィルタを用いず、帯域除去フィルタだけで実現すれば、事前の学習も必要としない。このようにして得られる、ガイダンス音声と同一周波数帯域だけを除去した音声は、発声者が入力する直接の音声とは音質  
50

的には異なるが、除去する周波数帯域を聴覚上問題のない帯域に限定すれば、音声入力装置より再生したときの音声波形は聴覚的には変化が少ない。また、この音声を音声認識装置で使用する場合には、認識に使用する標準パターン自体もこのような帯域除去フィルタを通過した音声によって構成しておけば、認識率の劣化を生じにくい結果が得られる。また、この方法では、ガイダンス音声自体の品質は低下するが、ガイダンス音声の音量を大きくすることにより、ガイダンス音声自体が聞き取りにくいといった現象は軽減できる。このように、ガイダンス音声と入力音声とで使用する周波数帯域を分離することにより、ガイダンス音声の重量があっても入力音声自体への影響がおよびにくい、ガイダンス音声付き音声入力装置が実現できる。このように第4の方法を改良することにより上述した欠点を解決することができる。

【0007】本発明は、上記に鑑みてなされたもので、その目的とするところは、ガイダンス音声に発声者の発声が重量されてもガイダンス音声の成分を含まない音声信号のみを取り出すことができるガイダンス音声付き音声入力装置を提供することにある。

【0008】

【課題を解決するための手段】上記目的を達成するため、本発明のガイダンス音声付き音声入力装置は、発声者に対して発声を促すガイダンス信号を生成する信号生成手段と、該信号生成手段の出力信号から特定の周波数帯域の信号を取り出す帯域通過フィルタと、該帯域通過フィルタの出力信号を発声者に受聴させるべく該出力信号を音声信号として発生する拡声器と、前記音声信号に応答して発声者から出力される音声を受信し、電気信号として出力する信号入力手段と、該信号入力手段の出力信号から前記特定の周波数帯域以外の信号を取り出す帯域除去フィルタとを有することを要旨とする。

【0009】

【作用】本発明のガイダンス音声付き音声入力装置では、ガイダンス信号を帯域通過フィルタに通すことにより特定の周波数帯域の信号に限定するとともに、このガイダンス音声にตอบสนองして発声者から出力される音声に対しては帯域除去フィルタを使用してガイダンス信号と同一の周波数帯域の信号を除去し、これによりガイダンス音声と発声者の音声信号とが時間的に重量しても、ガイダンス信号の成分を含まない音声信号を得ることができる。

【0010】

【実施例】以下、図面を用いて本発明の実施例を説明する。

【0011】図1は本発明の一実施例に係わるガイダンス音声付き音声入力装置の構成を示すブロック図である。同図において、ガイダンス信号源11は、発声者19に発声を促すガイダンス音声を出す信号源であり、例えば、テープレコーダーで録音した音声や、また例え

ば、電氣的に合成した音声である。帯域通過フィルタ12は、特定の周波数帯域の信号だけを通過させることを目的としたフィルタであり、その通過帯域は、例えば、1kHzから2kHzとする。また、その通過帯域は、一つの帯域に限る必要はない。増幅器13は、帯域通過フィルタ12を通過した後の信号を歪なく増幅するものである。スピーカ14は、増幅器13の出力信号を空間へ放出し発声者19へ受聴させるものである。マイクロホン15は、発声者19の音声を収音するマイクロホンである。帯域除去フィルタ16は、帯域通過フィルタで通過を目的とする周波数帯域の信号を除去し、また、その他の周波数帯域の信号について通過させることを目的としたフィルタであり、その通過帯域は、例えば、100Hzから1kHz、および、2kHzから5kHzとする。音声出力部17は、帯域除去フィルタ16を通過した音声信号を出力するものであり、例えば、電氣的な端子、テープレコーダー、磁気メモリなどである。スイッチ18は、発声者が本実施例の装置に対して動作を要求するためのスイッチであり、例えば、電氣的な接点、あるいは、光センサーなどである。

【0012】以下に図1の実施例の作用を説明する。

【0013】発声者19が本装置に対して音声入力するとき、まず、発声者はスイッチ18を操作し、装置の動作要求を与える。動作要求に対して、装置は、要求を受け入れたことを発声者19に確認させるために、ガイダンス信号源11より、「音声を入力して下さい」などの音声信号を生成する。この信号は、帯域通過フィルタ12により、1kHzから2kHzの特定の周波数帯域だけの信号となり、増幅器13を通じてスピーカ14より放出される。また、スイッチ18の動作要求に対して、マイクロホン15、帯域除去フィルタ16、および音声出力部17についても動作を開始し、発声者による音声の入力を待つ。入力を待つ時間は、例えば、数秒から数十秒間の有限な時間とする。ここで、ガイダンス信号がスピーカ14より放出されている間に、あるいは、放出後に、発声者19からの発声があると、この音声は、マイクロホン15より入力され、帯域除去フィルタ16を通過することにより1kHzから2kHzの特定の周波数帯域について除去された音声信号となり、音声出力部17より、ガイダンス信号帯域を含まない音声信号を出力する。

【0014】図2は本発明の他の実施例のブロック図であり、この実施例の音声入力装置は音声認識装置に適用されている。

【0015】図2において、ガイダンス信号源21は、図1のガイダンス信号源と同様に、発声者33に発声を促すガイダンス音声を出す信号源であるが、また、例えば、後述するラベル出力部31の出力結果をもとに、認識結果のラベルを発声者33に確認させるために出力する機能を有する。帯域通過フィルタ22、増幅器23、

5

スピーカ24、マイクロホン25、帯域除去フィルタ26は、それぞれ、図1における帯域通過フィルタ12、増幅器13、スピーカ14、マイクロホン15、帯域除去フィルタ16と同様の機能を持つものである。音声分析部27は、帯域除去フィルタ26より出力された音声信号からスペクトル領域の特徴を表すパラメータを算出するものであり、例えば、短時間LPC係数分析などを用いる。特徴記憶部28は、音声分析部27で算出されたパラメータの時系列を時間区間単位で記憶し、ラベルを付与して、音声認識で使用する標準パターンを構成するものであり、その時間区間は、例えば、本実施例を単語音声認識装置として使用するならば、1単語区間について記憶するものとする。ラベル入力部29は、特徴記憶部28で蓄積した1時間区間毎について、例えば、発声者33の手入力により、または、例えば自動的な順序で、ラベルを付与し、特徴記憶部28に記憶させるものである。類似度計算部30は、ラベルが未知の音声として音声分析部27より入力されたパラメータ列と、特徴記憶部28に記憶される標準パターンとの間で類似度計算を行うものであり、その計算方法は、例えば、DPマッチングを使用する。ラベル出力部31は、類似度計算部30の結果よりラベルが未知の入力音声に対して最も類似度が高いと判定された標準パターンのラベルを出力するものであり、また、その出力は同様にガイダンス信号源21に与えられる。スイッチ32は、上記音声認識装置を動作させるためのスイッチである。

【0016】以下に、図2のブロック図の動作を説明する。

【0017】本実施例は、音声認識で使用する標準パターンを記憶する登録モードと、音声を入力して標準パターンとの類似度を計算し、最も類似度の高い標準パターンのラベルを出力する認識モードの2つの動作を行う。

【0018】最初に、登録モードの動作について説明する。

【0019】まず、発声者33はスイッチ32を認識モードへ操作し、認識装置の動作要求を与える。動作要求に対して、認識装置は、要求を受け入れたことを発声者33に確認させるために、ガイダンス信号源21より、例えば「認識単語を発声して下さい」などの、ガイダンス音声が発生する。この信号は帯域通過フィルタ22により1kHzから2kHzの特定の周波数帯域だけの信号となり、増幅器23を通じてスピーカ24より放出される。このとき同時に、マイクロホン25、帯域除去フィルタ26、および音声分析部27についても動作を開始し、発声者による音声の入力を待つ。入力を待つ時間は、例えば、スイッチ32の操作開始より数秒から数十秒以内とする。ここで、ガイダンス信号がスピーカ24より放出されている間に、あるいは、放出後に、発声者33からの発声があると、この音声は、マイクロ

6

ホン25より入力され、帯域除去フィルタ26を通過することにより1kHzから2kHzの特定の周波数帯域を除去された音声信号となり、音声分析部27においてガイダンス信号を含まない音声信号として分析される。このとき、音声分析部27では、音声区間のみを選択して分析結果を出力する。音声区間の検出方法は、例えば、音声分析部27の出力における短時間信号パワーを観測し、信号パワーがある閾値を一定時間以上越える区間を検出するようにして実現する。音声分析部27の出力は、認識に使用する標準パターンとして、ラベル入力部29からラベル付けが行われ、特徴記憶部28に記憶される。次に、認識モードの動作について説明する。

【0020】発声者33はスイッチ32を認識モードへ操作し、認識装置の動作要求を与える。動作要求に対して、認識装置は、要求を受け入れたことを発声者33に確認させるために、ガイダンス信号源21より、例えば「音声を入力して下さい」などの、ガイダンス音声が発生する。この信号は、登録モードの場合と同様に、帯域通過フィルタ22により1kHzから2kHzの特定の周波数帯域だけの信号となり、増幅器23を通じてスピーカ24より放出される。このとき同時に、マイクロホン25、帯域除去フィルタ26、および音声分析部27についても動作を開始し、発声者による音声の入力を待つ。入力を待つ時間は、例えば、スイッチ32の操作開始より数秒から数十秒以内とする。ここで、ガイダンス信号がスピーカ24より放出されている間に、あるいは、放出後に、発声者33からの発声があると、この音声は、登録モードの場合と同様に、マイクロホン25より入力され、帯域除去フィルタ26を経由して1kHzから2kHzの特定の周波数帯域を除去された音声信号となり、音声分析部27より音声区間のみを選択されて出力される。音声分析部27の出力は、ラベルが未知の入力音声パターンとして、特徴記憶部28に記憶された複数の標準パターンと類似度計算部30においてパターンマッチングされる。この結果、最も類似度の高いと判定された標準パターンのラベルがラベル出力部31より出力される。出力されたラベルは、例えば、ガイダンス信号源21へフィードバックして、発声者33への認識結果の確認に使用することも可能である。例えば、標準パターンとして登録した音声ラベルを、「すずき」、「たなか」、「さとう」、とすると、認識した標準パターンのラベルを「すずき」としたとき、「認識単語は「すずき」です。よろしいですか」などの合成音声をガイダンス信号源21から発生して発声者33に聴かせて確認させ、また、これに対する「はい」、「いいえ」などの確認の音声を発声者33に入力させることができる。

【0021】本実施例の音声認識装置において、ガイダンス音声信号の重畳に対して認識性能の低下が生じにくいことを確認するために、実施例を模擬した実験を行っ

7

た。実験は、ガイダンス音声を重畳させた、成人男性5名が発声した日本の都市名100単語を認識するものである。ガイダンス音声は、20歳代の成人女性が「コマンド名を発声して下さい」と発声したものであり、これを先の説明で述べた帯域通過フィルタ22と同等の1kHzから2kHzの周波数帯域だけを通過するフィルタに通し、発声者の原音声の短時間パワーと同レベルのパワーになるように、「コマンド名」の部分で重畳させた。このガイダンス重畳音声について、帯域除去フィルタでガイダンスを除去した場合と、除去しない場合との音声認識率の変化を見た。使用した帯域除去フィルタは、さきに説明した帯域除去フィルタ26と同等の1kHzから2kHzの周波数を除去するフィルタであり、ガイダンス重畳音声について、ガイダンスの除去をする場合は、このフィルタを通して、また、ガイダンスの除去をしない場合はこのフィルタを通さずに、16次のLPCケプストラム分析を行い、音声パターンを作成して認識実験を行った。なお、標準パターンは、標準パターンと入力音声パターンとがフィルタの通過の有無も含めて音響的に同一の条件で分析されるように設定して作成した。話者5名について平均した実験結果では、ガイダンス音声の重畳が無い場合に、認識率98.6%であった音声について、ガイダンス音声为重畳し、これを帯域除去フィルタで除去しなかった場合、認識率84.8%と約14%の低下を見た。これに対して、帯域除去フィルタで除去した音声では、認識率93.6%となり、認識率の低下を5%に抑えられることを確認した。このように、ガイダンス付き音声認識装置においては、帯域制限したガイダンス音声を使用した場合、ガイダンス音声の重畳があっても、これを帯域除去フィルタで除去することにより、認識率の劣化を少なく抑えることができる。

【0022】

8

【発明の効果】以上説明したように、本発明によれば、ガイダンス信号を帯域通過フィルタに通すことにより特定の周波数帯域の信号に限定するとともに、このガイダンス音声にตอบสนองして発声者から出力される音声に対しては帯域除去フィルタを使用してガイダンス信号と同一の周波数帯域の信号を除去し、これによりガイダンス音声と発声者の音声信号とが時間的に重畳しても、ガイダンス信号の成分を含まない音声信号を得ることができるので、ガイダンス発声用のスピーカと音声入力用のマイクロホンとを音響的に分離して配置する必要がなく、ハンズフリーの音声入力装置を構成することが容易であるとともに、ガイダンス信号が開始された時点から任意の時刻で発声を受け付けることができ、親和性のある音声入力装置を構成でき、また適応予測フィルタ等を使用した従来のガイダンス音声除去方法で必要としたフィルタの学習アルゴリズム、学習のためのガイダンス音声信号および学習時間を排除することができる。更に、本音声入力装置を音声認識装置に適用した場合に、音声にガイダンス音声为重畳したり、ガイダンス音声の除去不足であっても、認識率の低下を生じない。

【図面の簡単な説明】

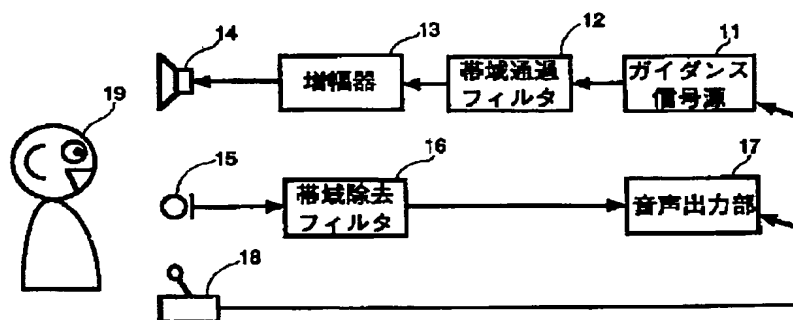
【図1】本発明の一実施例に係わるガイダンス音声付き音声入力装置の構成を示すブロック図である。

【図2】本発明の他の実施例に係わるガイダンス音声付き音声入力装置を音声認識装置に適用した場合の構成を示すブロック図である。

【符号の説明】

- 11 ガイダンス信号源
- 12 帯域通過フィルタ
- 14 スピーカ
- 15 マイクロホン
- 16 帯域除去フィルタ
- 17 音声出力部

【図1】



【図2】

